Subspace Properties of Randomized Network Coding

Mahdi Jafarisiavoshani, Christina Fragouli, Suhas Diggavi

Abstract— **Randomized network coding has network nodes randomly combine and exchange linear combinations of the source packets. A header appended to the packet, called coding vector, specifies the exact linear combination that each packet carries. The main contribution of this work[1] is to investigate properties of the subspaces spanned by the collected coding vectors in each network node. We use these properties to exhibit the relationship between the network topology and the subspaces collected at the nodes. This allows us to passively infer the network topology for a general class of graphs.**

I. INTRODUCTION

Consider a network $G = (V, E)$ where a source $S \in V$ has a set of n independent packets to distribute to a set of receivers using network coding techniques [1], and each packet is a sequence of symbols over a finite field $\mathbb{F}_q$. We can think of each source packet as corresponding to one dimension of an n-dimensional space over $\mathbb{F}_q$. We can thus associate with each packet one of the orthonormal basis vectors $\{e_1, \ldots, e_n\}$, where e_i is the n-dimensional vector with one at position i and zero elsewhere.

In randomized network coding, every node sends uniform at random linear combinations over $\mathbb{F}_q$ of its collected packets to its neighbors [2]. To enable decoding, an n-dimensional vector $\mathbb{F}_q^n$, called coding vector, is appended to each packet, to denote the expansion of the packet with respect to the n basis source packet vectors [3]. We say that node $i \in V$ at time t observes a subspace $\Pi_i(t) \subseteq \mathbb{F}_q^n$, if $\Pi_i(t)$ is the space spanned by the received coding vectors at node i up to time t. When $\dim(\Pi_i) = n$, node i has collected a basis of the n-dimensional space, and can decode the source information.

In this paper, we establish a relationship between the network topology and the subspaces observed by the nodes over time. We are interested in the conditions under which there exists a unique topology corresponding to a given set of $|V|$ observed subspaces. Although the focus of this paper is on understanding the connection between subspaces and network topology from a theoretical point of view, this work has also a variety of applications, for example on examine such applications in other work [4], [5].

Taking advantage of network coding for active network monitoring was proposed in [6], [7], where the focus was on link loss rate inference. Passive inference of link loss rates has also been proposed in [8]. The connection between subspaces and topological properties of the network is a novel contribution of this paper.

The paper is organized as follows. Section II describes our model; Section III investigates properties of randomly chosen subspaces and their evolution; Section IV presents results for tree topologies; Section V generalizes our work to arbitrary topologies; and finally Section VI concludes the paper with a short discussion.

II. NETWORK MODEL

Consider a network represented as a connected graph $G = (V, E)$, with $|V|$ nodes and $|E|$ edges, and assume that each edge has integer (and positive) capacity. The connection between network topology and subspaces observed depends on the dissemination protocol. It can be further aided by the prior information about the network structure.

For example, we can consider a synchronous or an asynchronous network operation model.

- *Synchronous:* All nodes are synchronized and transmit to their neighbors according to a global clock tick (time-slot). At time t node i sends linear combinations from all vectors it has collected up time $t - 1$, chosen uniformly at random from $\Pi_i(t - 1)$. Once nodes start transmitting information, they keep transmitting until all receivers are able to decode.
- *Asynchronous:* Nodes transmit linear combinations at randomly and independently chosen time instants.

Moreover, we may assume that we have

- *Global information*: A central entity knows the subspaces that all $|V|$ nodes in the network have observed.
- *Local Information:* There is no such omniscient entity, and each node i only knows what it has received, its own subspace Π_i.

Finally we may have a *static view*, where we take a snapshot of the network at a given time instant t, or a *non-static view*, where we take several snapshots of the network and use the subspaces' evolution to infer topological information. In this paper we will

1) restrict our attention to synchronous operation (a similar approach can be used for the asynchronous case),
2) consider networks with a single source S injecting packets (similar techniques apply for multiple sources), and
3) assume global information. Global information is the maximum information we can hope to get from the coding vectors propagated through the network. We focus on local information in a companion paper [4].

Definitions and Notation

Let $\Pi_i(t)$ denote the subspace node i has collected up to time t. For simplicity of notation, we will drop t when not necessary and use Π_i. To compare subspaces Π_i and Π_j, we will denote

1. the dimension of each subspace as

$$d_i \triangleq \dim(\Pi_i), \quad \forall i,$$

2. the dimension of the intersection of two subspaces as

$$d_{ij} \triangleq \dim(\Pi_i \cap \Pi_j), \quad \forall i, j,$$

3. the dimension of the joint span of two subspaces as

$$D_{ij} \triangleq \dim(\Pi_i \cup \Pi_j) \quad \forall i, j,$$

where by union we mean the common span of Π_i, Π_j, *i.e.*, $\Pi_i \cup \Pi_j = \mathrm{span}\{\Pi_i, \Pi_j\}$. Note that $d_i + d_j = d_{ij} + D_{ij}$. For a set of nodes $\mathcal{U} = \{u_1, \ldots, u_m\}$, we will denote as $d_{\mathcal{U}} \triangleq \dim(\Pi_{u_1} \cup \ldots \cup \Pi_{u_m})$.

Initially, at time $t = 0$, the subspaces of all nodes (apart the source) are empty. We define the transition phase to be the time during which some of the nodes have started receiving packets, while other nodes have not.

Definition 1: The *transition phase threshold* τ of an algorithm, is the first time τ at which each edge of the network has been used at least once.

We say that a network is in *steady state phase* if $t > \tau$, and none of the receivers is able to decode the source packets. Throughout this paper, we require that the topology identification occurs at *any time* during the steady state phase of the network.

We will consider connected networks, where each node, apart from the source, has at least one node (parent) transmitting information to it. If node i has p_i parents $u_1, \ldots, u_{p_i}$, we will denote with $\hat{\Pi}_{u_j}^{(i)}(t)$ the subspace node i has received from parent u_j up to time t, and with $\hat{\pi}_{u_j}^{(i)}(t)$ the subspace node i receives from parent u_j at exactly time t. Thus, $\hat{\Pi}_{u_j}^{(i)}(t) = \hat{\Pi}_{u_j}^{(i)}(t-1) \cup \hat{\pi}_{u_j}^{(i)}(t)$, and $\Pi_i(t) = \cup_{j=1}^{p_i} \hat{\Pi}_{u_j}^{(i)}(t)$.

III. PROPERTIES OF SUBSPACES

A. Subspaces in General Position

The intuition why looking at subspaces can enable us to distinguish between topologies, is that randomly chosen subspaces over a large enough finite field tend to be in *general position*. We say that a set of m vectors are in general position over an n-dimensional space $\mathbb{F}_q^n$ if every $k \leq n$ of these vectors are linearly independent [9]. The notion of general position extends over subspaces: two subspaces are said to be in general position if they are "as far away as possible", *i.e.*, given their dimension, they have the smallest possible intersection [9].

The following lemmas, that we are going to use in the remaining of the paper, prove such general position properties.

Lemma 1: Construct the subspaces Π_1 and Π_2 of the n-dimensional space $\mathbb{F}_q^n$, by choosing $m_1 \leq n$ and $m_2 \leq n$ vectors respectively, uniformly at random from $\mathbb{F}_q^n$. Then with high probability[2]

1) the subspaces have the maximum dimension possible, i.e., $\Pr[d_1 = m_1] \approx 1$, $\Pr[d_2 = m_2] \approx 1$, and
2) the intersection of the subspaces is the minimum possible $Pr[d_{12} = \max\{d_1 + d_2 - n, 0\}] \approx 1$.

Proof: For the first claim, and $i = 1, 2$, it holds that

$$\Pr(\dim(\Pi_i) = m_i) = \prod_{j=0}^{m_i-1}(1 - q^{j-n})$$

$$\geq \left(1 - \sum_{j=0}^{m_i-1} q^{j-n}\right) \approx 1$$

(see also [2]). To prove the second claim, let $\Pi_1^{\perp}$ be the orthogonal complement to Π_1, *i.e.*, the unique subspace that has zero intersection with Π_1 and satisfies the property $\Pi_1 \cup \Pi_1^{\perp} = \mathbb{F}_q^n$. Let also $U = \{u_1, \ldots, u_{m_1}\}$ and $V = \{v_1, \ldots, v_{n-m_1}\}$ be sets of basis vectors for Π_1 and $\Pi_1^{\perp}$ respectively, where we used that, from (1), Π_1 has dimension m_1. We can then expand the m_2 randomly selected vectors $W = \{w_1, \ldots, w_{m_2}\}$ that generate Π_2 as

$$w_i = \sum_{j=1}^{m_1} \alpha_j^{(i)} u_j + \sum_{j=m_1+1}^{n} \alpha_j^{(i)} v_{j-m_1}, \quad i = 1, \ldots, m_1.$$

Let A be the $n \times m_2$ matrix with columns the vectors $\alpha^{(j)}$, and denote by $\widetilde{U}_{m_1 \times m_2}$, $\widetilde{V}_{(n-m_1) \times m_2}$ the sub matrices of A collecting the coefficients with the respect to U and V:

$$A = \begin{bmatrix} | & & | \\ \alpha^{(1)} & \cdots & \alpha^{(m_2)} \\ | & & | \end{bmatrix} = \left[\begin{array}{c} \widetilde{U}_{m_1 \times m_2} \\ \hline \widetilde{V}_{(n-m_1) \times m_2} \end{array} \right].$$

To calculate $d_{12} = \dim(\Pi_1 \cap \Pi_2)$, note that each $n \times 1$ vector in $\Pi_1 \cap \Pi_2$, (i) since it belongs in Π_2 it can be expanded with respect to W using a unique $m_2 \times 1$ coefficient vector α, and (ii) since it belongs in Π_1 the vector α satisfies the equation

$$\left[\begin{array}{c} \widetilde{U}_{m_1 \times m_2} \\ \hline \widetilde{V}_{(n-m_1) \times m_2} \end{array} \right] \cdot \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{m_2} \end{bmatrix} = \left[\begin{array}{c} b_1 \\ \vdots \\ b_{m_1} \\ \hline 0_{(n-m_1) \times 1} \end{array} \right].$$

Thus α belongs in the kernel (null space) of the matrix $\widetilde{V}_{(n-m_1) \times m_2}$. For $q \gg 1$ this matrix is full rank with high probability. As a result,

$$\dim(\mathrm{Kernel}(\widetilde{V}_{(n-m_1) \times m_2})) = m_2 - \mathrm{Rank}(\widetilde{V}_{(n-m_1) \times m_2})$$
$$\approx m_2 - \min(m_2, n - m_1)$$
$$= \max\{m_1 + m_2 - n, 0\}. \blacksquare$$

[2] The $\approx$ notation means that this is true with probability 1 as $q \to \infty$.

Under the conditions of Lemma 1, if for example $d_1 + d_2 < n$, the subspaces are disjoint, while if $d_1 < n$ and $d_2 < n$ they are distinct and differ in at least one dimension.

Lemma 2: Let Π_i and Π_j be subspaces of $\mathbb{F}_q^n$ with dimension d_i and d_j respectively and intersection of dimension d_{ij}. Construct Π_i' by choosing m vectors from Π_i uniformly at random. Then $\Pr(\Pi_i' \subset \Pi_j) \approx 0$, if $\Pi_i \nsubseteq \Pi_j$.

Proof: The probability that *all* m vectors are in the intersection is

$$\Pr(\Pi_i' \subset \Pi_j) = \left(\frac{q^{d_{ij}}}{q^{d_i}} \right)^m = q^{(d_{ij} - d_i)m},$$

which is of order $O(1/q)$ provided that $d_{ij} < d_i$. ∎

B. Min-cut and Innovative Information

The following theorem considers the rate at which the dimension of the subspace a node observes increases.

Theorem 1: Consider a synchronous network operation, and randomized network coding over a field $\mathbb{F}_q$. Then each node i receives innovative packets from the source at a rate that is upper-bounded by min-cut(S, i). It receives innovative packets at a rate exactly equal to min-cut(S, i), if the network is in the steady state and the field size q is sufficiently large.

Proof: The proof uses the algebraic approach in [10] to express the transfer matrix between each node i and the source, and the Schwartz-Zippel lemma to upper bound the probability that randomly chosen values for the linear combinations lead to a transfer matrix with rank equal to min-cut(S, i). The detailed proof is given in [5]. ∎

IV. TREE TOPOLOGIES

Let $G = (V, E)$ be a network that is a directed tree of depth[3] $\mathcal{D}$, rooted at the source node S. We will present (i) necessary and sufficient conditions under which the tree topology can be uniquely identified, and (ii) given that these conditions are satisfied, algorithms that allow us to do so.

We first consider trees where each edge has the same capacity c, and thus the min-cut from the source to each node of the tree equals c. We then briefly discuss the case of undirected trees. Finally we examine the case where edges have different capacities, and thus nodes may have different min-cuts from the source.

A. Common Min-Cut

Assume that each edge of the tree has capacity c, and consider the following dissemination algorithm, also summarized[4] in Algorithm IV.1. Each node i waits until its subspace dimension becomes m_i, *i.e.*, $d_i \geq m_i$ (for this section we will use a common value $m_i = m$). It then starts transmitting to each of its children c random linear combinations per timeslot.

[3] The depth of a tree is the length of the longest path between the root and a leaf of the tree.

[4] Though the Algorithm IV.1 is introduced for trees, it will also be used for general topologies in Section V.

Algorithm IV.1: INPUT$(G = (V, E), S, \{m_i\}, n)$

for each $i \in V \setminus \{S\}$
  do $\Pi_i(0) = \emptyset, d_i(0) = 0$
$t \leftarrow 0$
while $\min_i d_i(t) < n$
  do $\begin{cases} \textbf{for each } i \in V \\ \textbf{if } d_i(t) \geq m_i \\ \quad \textbf{then } \text{node } i \text{ transmits from } \Pi_i(t) \\ t \leftarrow t + 1 \\ \textbf{for each } i \in V \text{ update } \Pi_i(t), \ d_i(t) \end{cases}$

The following theorem presents necessary and sufficient conditions that enable us to identify the network topology using a single snapshot of all node's subspaces at a time t.

Theorem 2: Consider a tree of depth $\mathcal{D}$ where each edge has capacity c, and the dissemination algorithm in IV.1. A static global view of the network at time t, with $(\mathcal{D} - 1)m < t < \frac{n}{c}$, allows to uniquely determine the tree structure, if and only if

$$c + 1 \leq m. \tag{1}$$

Proof: The proof is based on the following simple observations. In a tree there exist a unique path $\mathcal{P} = \{S, i_1, \ldots, i_l, i\}$ from source S to node i. Clearly, in steady-state, for the nodes along the path it holds that

$$\Pi_i \subset \Pi_{i_l} \subset \cdots \subset \Pi_{i_1} \subset \mathbb{F}_q^n = \Pi_S. \tag{2}$$

The conditions on t ensure that the network is in steady-state, *i.e.*, all nodes have a non-empty subspace and no node's subspace (apart the source) equals the complete n-dimensional space.

Thus to identify the topology of the tree it is sufficient to show that $\Pi_i \nsubseteq \Pi_j$ for any j that is not in $\mathcal{P}$. But this is what the condition in (1) ensures. Indeed, consider a node $u \in V$ in the tree that has k children $u_1, \ldots, u_k$. If (1) holds, from Lemma 1, then $\Pi_{u_i} \neq \Pi_{u_j}$ for all i, j if and only if $m \geq c+1$ and $q \gg 1$. ∎

Thus the condition (1) on m ensures that the subspaces of all nodes in the tree are distinct during the steady-state phase. Obviously, if two nodes observe exactly the same subspace at time t, we can never distinguish between them; ensuring distinct subspaces is clearly necessary for identifiability.

The simple network in Fig. 1 can help us better understand why the conditions on m in Theorem 2 are both necessary and sufficient. Assume that the edges have unit capacity ($c = 1$). At time $t = 1$, node A receives a vector y_1 from the source S. If node A starts transmitting to nodes B and C at time $t = 2$, then nodes B and C will both receive the same vector y_1, *i.e.*, $\Pi_B(2) = \Pi_C(2) = \text{span}\{y_1\}$. In fact, at all subsequent times, we will have that $\Pi_B(t) = \Pi_C(t) = \Pi_A(t - 1)$. If instead, node A waits to collect $c + 1 = 2$ vectors, say y_1 and y_2, before starting transmission to nodes B and C, then it will hold that $\Pi_B(t) \neq \Pi_C(t)$, for $2 \leq t \leq n + 1$.

Assume now that Theorem 2 holds. To determine the tree structure, it is sufficient to determine the unique parent each node has. From the previous arguments, the parent of node i
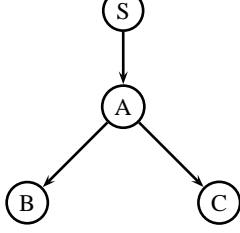
Fig. 1. Directed tree with four nodes rooted at the source S.

is the unique node j such that Π_j is the minimum dimension subspace that contains Π_i. Then, the parent of node i is the node j such that, $j = \operatorname{argmin}_{k:d_{ik}=d_i} d_k$. Note that to determine the tree topology, we do not need to know exactly which are the node subspaces, but only two "sufficient statistics": the dimension of each subspace $d_i = \dim(\Pi_i)$, $\forall i$, and the dimension of the intersection of every two subspaces $d_{ij} = \dim(\Pi_i \cap \Pi_j)$, $\forall i,j$, as described in Algorithm IV.2, assuming that the conditions of Theorem 2 hold.

Algorithm IV.2: TREE($\{d_i\}, \{d_{ij}\}$)

for each $i \in V$

do $\begin{cases} \textbf{if } d_i = n, i \leftarrow S \\ \quad \textbf{else} \text{ node } i \text{ has parent the node } j \text{ with} \\ \quad j = \operatorname{argmin}_{k:d_{ik}=d_i} d_k \end{cases}$

B. Directed v.s. Undirected Network

In a tree with a single source, since new information can only flow from the source to each node along a single path, whether the network is directed or undirected makes no difference. In other words, from condition (2), all vectors that a node will send to its predecessor will belong in the subspace the predecessor already has. Thus Theorem 2 still holds for undirected networks with a common min-cut.

C. Different Min-Cuts

Assume now that the edges of the tree have different capacities. As a result, potentially $\operatorname{mincut}(S,i) \neq \operatorname{mincut}(j,i)$, for some node j in the path $\mathcal{P}$ that connects node i to the source S. Note that, under Algorithm IV.1, for the subspaces of the nodes in the path between S and i, condition (2) still holds. However, it is possible that we cannot distinguish between nodes at same level with a common parent. For example, if in the network in Fig. 1, edge SA has unit capacity, while edge AB and AC have capacity two. In this case it is easy to see that there exists t_0 such that $\Pi_B(t) = \Pi_C(t) = \Pi_A(t-1)$, $\forall t \geq t_0$. Clearly in this case, we cannot distinguish between nodes B and C with this dissemination protocol.

V. GENERAL TOPOLOGIES

Consider now an arbitrary network topology, corresponding to a directed graph. An intuition we can get from examining tree structures is that, we can distinguish between two topologies provided all node subspaces are distinct. The following

theorem[5] claims that this is in fact a sufficient condition for topology identifiability over general graphs.

Theorem 3: In a synchronous network employing randomized network coding over $\mathbb{F}_q$, a sufficient condition to uniquely identify the topology with high probability as $q \gg 1$, is that

$$\Pi_i(t) \neq \Pi_j(t) \quad \forall \, i,j \in V, \quad i \neq j, \qquad (3)$$

for some time t. We can achieve this by collecting global information at times t and $t+1$, *i.e.*, two consecutive static views of the network.

Proof: Assume node i has the p_i parents $P(i) = \{u_1, \ldots, u_{p_i}\}$. Let $\hat{\Pi}_{u_1}^{(i)}(t), \ldots, \hat{\Pi}_{u_{p_i}}^{(i)}(t)$ denote the subspaces node i has received from its parents up to time t, where $\Pi_i(t) = \cup_{j=1}^{p_i} \hat{\Pi}_{u_j}^{(i)}(t)$. From construction it is clear that $\hat{\Pi}_{u_j}^{(i)}(t+1) \subseteq \Pi_{u_j}(t)$.

To identify the network topology, it is sufficient to decide which node $v \in V$ is the parent that sent the subspace $\hat{\Pi}_{u_j}^{(i)}(t)$ to node i for each j, and thus find the p_i parents of node i. We claim that, provided (3) holds, node i has as parent the node v which at time t has the smallest dimension subspace containing $\hat{\Pi}_{u_j}^{(i)}(t+1)$. Thus we can uniquely identify the network topology, by two static views, at times t and $t+1$, as Algorithm V.1 describes.

Indeed, let $\hat{\pi}_{u_j}^{(i)}(t)$ denote the subspace that node i receives from parent u_j at exactly time t, that is, $\hat{\Pi}_{u_j}^{(i)}(t+1) = \hat{\Pi}_{u_j}^{(i)}(t) \cup \hat{\pi}_{u_j}^{(i)}(t+1)$.

• If $\hat{\pi}_{u_j}^{(i)}(t+1) \not\subseteq \Pi_v(t)$ for all $v \in V \backslash \{u_j\}$, clearly $\hat{\Pi}_{u_j}^{(i)}(t+1) \not\subseteq \Pi_v(t)$ for all $v \in V \backslash \{u_j\}$, and we are done.

• Assume now there exist two nodes j and k such that $\hat{\Pi}_{u_j}^{(i)} \subseteq \Pi_j \subset \Pi_k$. From Lemma 2, node i cannot be a child of node k, because then we would have that $\hat{\pi}_{u_j}^{(i)} \not\subseteq \Pi_j$, and as a result, $\hat{\Pi}_{u_j}^{(i)} \not\subseteq \Pi_j$. Thus it can only be a child of node j. ∎

Note that to identify the network topology, we need to know, for all nodes i, the dimension of their observed subspaces at time t, the dimension $\hat{d}_{u_j^{(i)}} \triangleq \dim(\hat{\Pi}_{u_j}^{(i)}(t+1))$ for all parents j of node i, and the dimension of the intersection of $\hat{\Pi}_{u_j}^{(i)}(t+1)$ with all $\Pi_k(t)$, denoted as $\hat{d}_{ku_j^{(i)}} \triangleq \dim(\hat{\Pi}_{u_j}^{(i)}(t+1) \cap \Pi_k(t))$. Algorithm V.1 uses this information to infer the topology.

Algorithm V.1: GEN($\{d_i(t)\}, \{\hat{d}_{u_j^{(i)}}\}, \{\hat{d}_{ku_j^{(i)}}\}$)

for each $i \in V$

do $\begin{cases} \textbf{if } d_i = n, i \leftarrow S \\ \quad \textbf{else} \text{ node } i \text{ has parent the node } j \text{ with} \\ \quad j = \operatorname{argmin}_{k:\hat{d}_{ku_j^{(i)}}=\hat{d}_{u_j^{(i)}}} d_k(t) \end{cases}$

The sufficient conditions in (3), Theorem 3, may or may not hold, depending on the network topology and the information

---

[5]Note that if we identify the parents of each node, we know the graph topology.

dissemination protocol. Next, we will investigate under what conditions there exist values $\{m_i\}$ for the simple dissemination algorithm IV.1 so that (3) holds, and the network topology is identifiable.

Lemma 3: Consider two arbitrary nodes i and j, where $P(i) = \{u_1, \ldots, u_{p_i}\}$ and $P(j) = \{v_1, \ldots, v_{p_j}\}$ are the parents of i and j respectively. Let $\Pi_{P(i)}(t-1) = \cup_{l=1}^{p_i} \Pi_{u_l}(t-1)$ and $\Pi_{P(j)}(t-1) = \cup_{l=1}^{p_j} \Pi_{v_l}(t-1)$. The condition $\Pi_{P(i)}(t-1) \neq \Pi_{P(j)}(t-1)$ is sufficient to guarantee that $\Pi_i(t) \neq \Pi_j(t)$.

Proof: Let us assume that $\Pi_i(t) = \Pi_j(t) = \Pi$. This implies that if $\pi_i(t)$ and $\pi_j(t)$ are subspaces collected at time t then,

$$\pi_i(t) \cup \Pi_i(t-1) = \pi_j(t) \cup \Pi_j(t-1) = \Pi.$$

From construction, $\Pi_i(t-1) \subseteq \Pi_{P(i)}(t-1)$ and $\pi_i(t) \subseteq \Pi_{P(i)}(t-1)$ so we have $\Pi \subseteq \Pi_{P(i)}(t-1)$. The same is true for node j, $\Pi \subseteq \Pi_{P(j)}(t-1)$.

On the other hand, using Lemma 2, since we randomly chose $\pi_i(t)$ from $\Pi_{P(i)}(t-1)$ and since $\pi_i(t)$ is a subspace of Π, we should have that $\Pi_{P(i)}(t-1) \subseteq \Pi$, and similarly that $\Pi_{P(j)}(t-1) \subseteq \Pi$. We conclude that

$$\Pi_{P(i)}(t-1) = \Pi_{P(j)}(t-1) = \Pi,$$

which gives us the result. ■

Now consider the parents of nodes i and j as supernodes $P(i)$ and $P(j)$. Using a similar argument we can conclude that the parents of $P(i)$ and $P(j)$, denoted as $P^2(i)$ and $P^2(j)$, satisfy

$$\Pi_{P^2(i)}(t-2) = \Pi_{P^2(j)}(t-2) = \Pi,$$

where $\Pi_i(t) = \Pi_j(t) = \Pi$, and $\dim(\Pi) = d < n$. Continuing this procedure, and including at least one new node in the set of parents at each step, we will at some step ℓ, either have $P^\ell(i)$ include the source node S, which leads to a contradiction since the dimension of the subspace $\Pi_{P^\ell(i)}(t-\ell)$ is $d < n$ (similarly if $P^\ell(j)$ includes the source S), or that $P^\ell(i) = P^\ell(j)$. To resolve this last case, we evoke the following theorem.

Theorem 4: Suppose two arbitrary nodes i and j have a common set of parents $P^\ell = P^\ell(i) = P^\ell(j)$ at a level ℓ. The following conditions are sufficient to let exist some $\{m_i\}$ for the Algorithm IV.1 such that (3) will be satisfied[6]:

$$\hat{c}_i = \text{min-cut}(P^\ell, i) \leq \text{min-cut}(S, P^\ell) = c_p,$$
$$\hat{c}_j = \text{min-cut}(P^\ell, j) \leq \text{min-cut}(S, P^\ell) = c_p.$$

Proof: Let us assume that t_0 is the first time that $\dim(\Pi_{P^\ell}) \geq c_p + 1$ and the time after which P^ℓ receives innovative packets at a rate of c_p. Assume that P^ℓ starts transmission after t_0. For t_1 time slots later we can write

$$\dim(\Pi_{P^\ell}(t_0 + t_1)) \geq t_1 c_p + c_p + 1.$$

For node i we can also write

$$\dim(\Pi_i(t_0 + t_1 + l)) \leq (t_1 + 1)\hat{c}_i \leq t_1 c_p + c_p.$$

[6]Note that if $c_i = \text{min-cut}(S, i)$, $c_i = \min\{\hat{c}_i, c_p\}$.

The same inequality holds for the dimension of $\Pi_j(t_0 + t_1 + l)$. Thus for $t - l > t_0$ we cannot have $\Pi_{P^\ell}(t-l) = \Pi_i(t)$ and $\Pi_{P^\ell}(t-l) = \Pi_j(t)$. Using Lemma 3 we are done. ■

Intuitively, what the previous theorems tell us is that, if for a node i there exists a path that does not belong in any cut between the source and another node j, then nodes i and j will definitely have distinct subspaces. The only case where nodes i and j may have the same subspace is, if they have a common set of parents, a common cut. Even then, they would need both of them to receive all the innovative information that flows through the common cut at the same time. Note that the condition of Theorem 4 are also necessary for identifiability for the special case of tree topologies, such as the topology in Fig. 1. We can develop dissemination techniques for general topologies that satisfy the sufficient conditions given in Theorem 4 by using a decentralized rate control strategy [5]. This can be done with almost no affect on the dissemination rate.

VI. CONCLUSIONS

In this paper we have shown that (for a class of graphs) one could design network coding algorithms which reveal topological structure of the graph while not affecting the dissemination rates. This connection between subspaces of network coded packets and network properties could be useful in other contexts as well. We have only considered the case where the identifiability occurs at *any time* during the steady state, and sufficient conditions for this. If one relaxes this or has further prior information about the network topology, one could also design other schemes [5].

REFERENCES

[1] R. Ahlswede, N. Cai, S-Y. R. Li, and R. W. Yeung, "Network information flow", *IEEE Trans. Inform. Theory,* vol. 46, pp. 1204–1216, July 2000.
[2] T. Ho, R. Kötter, M. Médard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast", *IEEE Trans. Imform. Theory,* vol. 52, pp. 4413-4430, October 2006.
[3] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", *Allerton,* Monticello, IL, October 2003.
[4] M. Jafarisiavoshani, C. Fragouli, S. Diggavi, and C. Gkantsidis, "Bottleneck discovery and overlay management in network coded peer-to-peer systems", *EPFL technical report, June 2007.*
[5] M. Jafarisiavoshani, MS thesis, EPFL, 2007.
[6] C. Fragouli and A. Markopoulou, "A network coding approach to overlay network monitoring", *Allerton*, Oct. 2005.
[7] C. Fragouli, A. Markopoulou, and S. Diggavi, "Active topology inference using network coding", *Allerton*, Oct. 2006.
[8] T. Ho, B. Leong, Y. Chang, Y. Wen and R. Kötter, "Network monitoring in multicast networks using network coding", *in International Symposium on Information Theory (ISIT)*, June 2005.
[9] L. Babai and P. Frankl, "Linear Algebra Methods in Combinatorics", preliminary version, University of Chicago.
[10] R. Kötter and M. Médard, "Beyond routing: an algebraic approach to network coding", *IEEE/ACM Trans. Networking,* vol. 11, pp. 782–796, 2003.